BIAS-VARIANCE TRADE-OFF IN MACHINE LEARNING

→ The prediction error of any machine learning algorithm can be broken down into three parts:

i) Bias error ii) Variance error iii) Irreducible error.

→ Irreducible error cannot be reduced regardless of what algorithm is used. It is the error introduced from the choosen framing of the problem and may be caused by the factors like unknown variables that influence the mapping of the input variables to the output variable.

→ Bias error,

- Bias are the simplifying assumptions made by a model to make the target function easier to learn.
- Low bias → Suggest less assumption about the form of target function
  Example - Decision tree, k-Nearest neighbour, Support vector machine.
- High bias → Suggest more assumption about the form of target function
  Example - Linear Regression, LDA and logistic Regression.

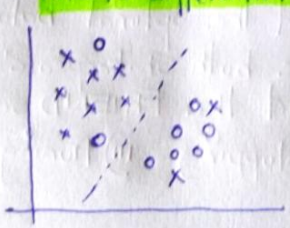
→ Variance error,

- Variance is the amount that estimates of the target function will change if different training data was used.
- Low variance → Suggests small changes to estimate of the target function with changes to the training dataset. For example linear regression, LDA (linear discriminant analysis) and logistic regression
- High variance → Suggests larger changes to estimate of the target function with changes to the training dataset. For example - Decision tree, k-NN and Support Vector Machines.

→ Goal of any supervised ML is to acheive low bias and low variance. But in general we can see linear ML algorithm often have a high bias but low variance and Non linear ML algorithm often have a low bias but high variance.

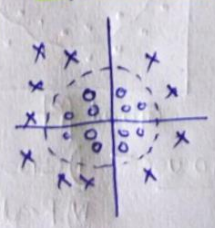
→ There is no escaping the relationship between bias and variance in ML.
  → Increasing the bias will decrease the variance
  → Increasing the variance will decrease the bias.

→ Linear data → In this we can draw a line to differentiate between data classes



Algorithm in linear are →
i) Linear Regression
ii) Logistic Regression.
iii) LDA.

→ Non linear data → Here we cannot seperate two class by straight line.



Algorithm in non linear
i) Classification & regression tree
ii) Naive bayes
iii) KNN
iv) SVM

Example of Bias Variance Tradeoff →

→ I wanted to learn English. I had no prior knowledge of the language but I heard the greatest English writer, William Shakeshpeare.

→ I locked myself in a library and memorized his work. After few weeks of study, I started speaking English like "Dear gentlewomen, How fares our gracious lady?". I got a reply "Are you crazy?".

→ I just committed basic mistake of "Overfitting on training data".

→ How Overfitting? The model I want to build was a representation of how to communicate using the English language. My training data was the entire work of shakespeare and testing set was Country speaking English. If I measure performance in terms of social acceptance, then my model fail to generalize or translate to testing data.

→ Variance is how much a model changes in response to training data. As we are simply memorizing the training set, our model has high variance, it is highly dependent on training data.

→ When a model with high variance is applied on a new testing set, it cannot perform well because all it is lost without training data.

→ Bias is the flip side of variance as it represents the strength of my assumptions I made about the data.

→ To summarize so far: bias refers to how much I ignored the data and variance refers to how much dependent my model is on the data.

→ So overfits means, high variance low bias. Underfitting means low variance, high bias.

→ Underfitting is instead of following the training data too closely, a model that underfits, ignores the lesson learnt for training data and fails to learn the underlying relationship between inputs and outputs.

→ Learning from previous failed attempt this time I take few assumption. I switched my training set and watch all episodes of "FRIEND" to teach myself english.

→ After training I again tested myself, this time fair but failed. Because while I know some English and can comprehend a limited number of sentences, I failed to learn fundamental structure of language because of my bias about teing data Model does not suffered from high variance but suffered from high bias (underfit)

Solution → Use Validation sets
Find the optimal balance.
Use both Shakespeare's work and the FRIEND series.

Common Methods to use → i) Cross Validation ii) Early Stopping iii) Pruning iv) Regularization.